

Combinatorial and approximative analyses in a spatially random division process

Yukio Hayashi ^a, Takayuki Komaki ^a Yusuke Ide ^b
Takuya Machida ^c Norio Konno ^d

^a*Japan Advanced Institute of Science and Technology, Ishikawa 923-1292, Japan*

^b*Kanagawa University, Kanagawa, 221-8686, Japan*

^c*Meiji University, Kanagawa, 214-8571, Japan*

^d*Yokohama National University, Kanagawa, 240-8501, Japan*

Abstract

For a spatial characteristic, there exist commonly fat-tail frequency distributions of fragment-size and -mass of glass, areas enclosed by city roads, and pore size/volume in random packings. In order to give a new analytical approach for the distributions, we consider a simple model which constructs a fractal-like hierarchical network based on random divisions of rectangles. The stochastic process makes a Markov chain and corresponds to directional random walks with splitting into four particles. We derive a combinatorial analytical form and its continuous approximation for the distribution of rectangle areas, and numerically show a good fitting with the actual distribution in the averaging behavior of the divisions.

Key words: Random Division Process; Universal Spatial Characteristic; Hierarchical Structure; Self-Organization; Complex Systems

PACS: 05.90.+m, 02.50.-r, 02.60.-x, 05.65.+b

1 Introduction

Historically, a discussion for the origin of skew distribution that appears in sociological, biological, and economic phenomena goes back to the Simon's stochastic model [1]. At the beginning of this century, the evidences of fat-tail degree distribution have been also observed in many real networks [2] through computer analyses for large data. Some cases look like a power-law distribution while some other cases a lognormal distribution, it is difficult to discriminate these distributions in general. Since the tail in a lognormal distribution resemble power-law behavior, only the part may be observed. Moreover, the generation mechanisms of the distributions are not exclusive but intrinsically connected in preferential attachment, multiplicative process, and other models [3]. On a stochastic process of geometric Brownian motion, a model of double Pareto distribution generates a lognormal body and Pareto tail in the continuous distribution [4][5]. On other process of iterative division of cells, the frequencies experimentally follow such distributions of fragment-size and -mass of glass [6][7], areas enclosed by city roads [8][9][10], pore size/volume in random packings [11][12], and areas enclosed by edges in models of urban street patterns [13] and geographical networks [14]. In addition, the distribution in fragmentation of glass changes from a lognormal to a power-law-like according to low and high impacts [6], which determine the limitation of breakable sizes. It is worthwhile to study a mechanism in abstract models for generating the similar distributions in spite of different physical quantities and operations in a variety of research fields: socio-economic, material, computer, and physical sciences, even if we apart from the reproduce of realistic phenomena and do not insist on the detail process or the macroscopic properties exactly in broken fragmentation of glass.

For crack patterns, a random tessellation model has been proposed [15]. It is based on a stochastic point process, consisting of the division of a randomly chosen face (cell) according to its life-time by adding a random line segment. The distribution of tessellations is invariant for an appropriate rescaling, whose characteristic is called stable with respect to iteration (STIT). In addition, the length distribution of segments is analytically obtained in a classification of several types of segments [16][17]. However, the distribution of areas enclosed by segments is not derived. The adjustment of life-time for each cell is also not easily applicable even in the sophisticated mathematical model of STIT, when we consider a construction of network in a procedural manner such as load balancing in a territory of node.

Thus, one of the issues is a theoretical analysis for the distribution of areas in a fractal-like structure. The presence of hierarchy and scaling law is important [18] for understanding the universal mechanism to generate such a structure. Through a mathematical model, we focus on the spatial phenomena based on

dividing by four with a randomness for the simplicity with analogous structures to road networks. However, for the iterative division processes, only a few mathematical models are known. In random tessellations, although the cell-selection and cell-division rules are classified into equally-likely, area-weighted, perimeter-weighted, corner-weighted, and so on, the length distributions in one-dimension are merely analyzed for some simple rules [19]. In a quadtree model characterized as a typical road network, the shortest paths and maximum flow are analyzed [20], but the distribution of areas is not discussed.

On the other hand, we recently proposed *multi-scale quartered* (MSQ) networks based on a self-similar tiling by equilateral triangles or squares [21,22]. This model is constructed by iterative division of faces, and is also suitable for the analysis of depth distribution of layered areas in a framework of Markov chain [23]. Moreover, from an application point of view, the MSQ networks without hub nodes have several advantages of the strong robustness of connectivity against node removals by random failures and intentional attacks, the bounded short path as $t = 2$ -spanner [24], the efficient face routing by using only local information, and a scalable load balancing performed by the divisions of the node's territory for increasing communication or transportation requests. However, due to the self-similar tiling in the scalably growing MSQ networks, the position of a new node is restricted on the half-point of an edge of the chosen face, and the link length is proportional to $(\frac{1}{2})^H$ where H is the hierarchical depth number of divisions. The restriction is unnatural for many division processes in physical or social phenomena. Thus, we generalize the divisions of squares to ones of rectangles with any link lengths instead of the iterative halvings.

The organization of this paper is as follows. In Sec. 2, we introduce a generalization of MSQ network model. In Sec. 3, for the distribution of areas, we derive the exact solution on a combinatorial analysis. We point out that the behavior of divisions is equivalent to directional random walks with splitting into four particles. The representation of random walks with splitting gives us an inspiration for the combinatorial analysis. However, this approach is limited for application to very small networks. In Sec. 4, we consider a continuous approximation of the distribution of areas for large networks. We decompose the distribution function into two components of Poisson and gamma distributions, and numerically investigate the fitting of the mixture distribution by the two components with the actual distribution of areas in the divisions for generating a fractal-like network structure. In Sec. 5, we summarize these results.

2 Division process

We consider a two-dimensional $L \times L$ square, whose lattice points $(0, 0), (0, 1), \dots, (0, L), \dots, (L, 0), (L, 1), \dots, (L, L)$ in the x - y coordinates give the feasible setting positions of nodes. Initially, there exist only the outer square with four corner nodes and edges of the length L . The following model generalizes the MSQ network [21,22,23] from recursive divisions of squares to ones of rectangles. In the MSQ network based on a self-similar tiling by squares, the division is restricted at the half point of an edge, therefore we extend the division to that at a cross point of the vertical and horizontal segments on the square lattice. The case of $L \rightarrow \infty$ gives a general position for the division point. Varying the value of L controls the limitation of divisible size relatively, as similar to low and high impacts in the fragmentation of glass [6].

The proposed network is iteratively constructed for a given L as follows. At each time step, a rectangle is chosen uniformly at random (u.a.r), and it is divided into four smaller rectangles. Then, the smaller rectangle with an area $x \times y$ (x, y denote the two edge lengths) is generated from the chosen rectangle with an area $x' \times y'$. Simultaneously, rectangles with the areas $(x' - x) \times y$, $x \times (y' - y)$, and $(x' - x) \times (y' - y)$ are generated. Here, two division axes are chosen u.a.r from the horizontal and vertical segments of an $L \times L$ lattice (see the left of Fig. 1). In other words, each edge length $x, y \in Z_+ = \{1, 2, \dots\}$ is randomly chosen as a positive integer in $x+1 \leq x' \leq L$ and $y+1 \leq y' \leq L$. The stochastic network generation makes a Markov chain. The state is represented by a vector $(n_{11}, \dots, n_{xy}, \dots, n_{LL})$, where n_{xy} denotes the number of rectangle with the area $x \times y$ (The Markov chain is degenerately simplified by ignoring the difference of areas in subsection 4.1 for discussing the distribution of layers defined by the depth of divisions.). The stochastic process is characterized by the fact that the transition probability to divide a rectangle with the area $x \times y$ is not fixed but proportional to n_{xy} because of the uniformly at random selection of a face. In other words, the probability depends on a sequence of chosen rectangles during the transition until a final absorbing state for the indivisible width $x = 1$ or $y = 1$. We remark that the minimum edge length $x = 1$ or $y = 1$ bounds the number of the states finitely, while the MSQ networks [23] have the infinite states without a limitation of the subdivision. The scaling relation of the maximum iteration time is numerically obtained as $T_{max} \sim L^{1.91}$ in the averaging of absorbing states. Although this paper discusses a simple case of uniformly random selections of rectangles and of division points in order to deduce the distribution of areas, the selections can be extended to other cases, e.g., according to a given population in a territory of node for real statistical data. Note that the above process is different from the Galton-Watson type branching process with a time-independent probability for generating offspring [25].

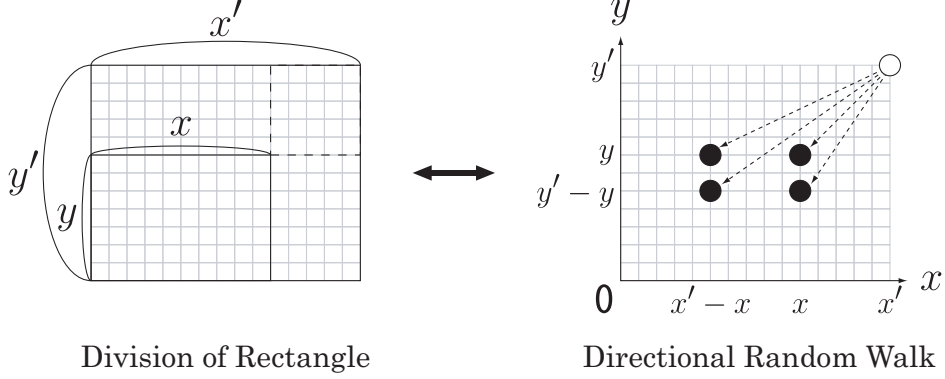


Fig. 1. Correspondence between the division of a rectangle and the directional random walks with splitting. Vertical and horizontal thin lines at even intervals are the segments of an $L \times L$ lattice.

3 Combinatorial analysis

As shown in Fig. 1, the recursive generation process can be regarded as directional random walks of particle with splitting into four copies in the framework of two-dimensional cellular automaton (CA), when a pair (x, y) of edge lengths of a rectangle is corresponded to the position of particle in the x - y coordinates. A particle is randomly chosen at a time step, and moves toward smaller coordinate values from (x', y') to (x, y) , where $x < x'$ and $y < y'$, until reaching the boundary at $x = 1$ or $y = 1$. We emphasize that this type of CA with splitting differs from the asymmetric simple exclusion process (ASEP) [26] and the contact (or voter) model [25], since there are no spatial exclusions and no interaction between any particles.

This representation of random walks with splitting is inspirational for deriving a combinatorial analytic form of the distribution of areas. We consider the number n_{xy} of particles at (x, y) , equivalently the number of rectangles with the area $x \times y$. Remember that x and y are integers. The average behavior is described by the following system of difference equations for $2 \leq x, y \leq L - 1$ with the sum by taking over the integers $x + 1 \leq x' \leq L$ and $y + 1 \leq y' \leq L$,

$$\Delta n_{xy} = -p_{xy} + \sum_{x', y'} \frac{4p_{x'y'}}{(x' - 1)(y' - 1)}, \quad (1)$$

where Δn_{xy} is the average difference of n_{xy} in one step, and $p_{xy} \stackrel{\text{def}}{=} n_{xy} / \sum_{x'' > 1, y'' > 1} n_{x''y''}$ is the existing probability of a particle at (x, y) . The factor 4 in the numerator of right-hand side of Eq.(1) is due to feasible positions of the $x \times y$ at left/right and upper/lower corners in the division of $x' \times y'$. The denominator is the combination number for the relative positions of emanating particles in the intervals $[1, x' - 1]$ and $[1, y' - 1]$, and is equivalent to the number for

selecting two axes in the division of rectangle with the area $x' \times y'$.

From $\Delta n_{xy} = 0$ in Eq.(1), we derive

$$\begin{aligned} p_{L-1L-1} &= \frac{4p_{LL}}{(L-1)^2}, \\ p_{xL-1} = p_{L-1y} &= \frac{4p_{LL}}{(L-1)^2}, \quad x > 1, y > 1, \\ p_{L-2L-2} &= \left(1 + \frac{4}{(L-2)^2}\right) \frac{4p_{LL}}{(L-1)^2}. \end{aligned}$$

In general, we obtain the solution by applying the above in decreasing order of x and y recursively.

$$p_{xy} = \left\{1 + \sum_{\mathcal{P}} \left(\prod_{i=1}^l \frac{4}{(x_i-1)(y_i-1)} \right) \right\} \frac{4p_{LL}}{(L-1)^2}, \quad (2)$$

where $\sum_{\mathcal{P}}$ denotes the sum for a set of paths through the points $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$, with $x_i, y_i \in \mathbb{Z}_+$, $x < x_1 < x_2 < \dots < x_i < \dots < x_l \leq L-1$, and $y < y_1 < y_2 < \dots < y_i < \dots < y_l \leq L-1$ in all combinations of $l = 1, 2, \dots, \min\{L-1-x, L-1-y\}$.

By substituting the solution p_{xy} of Eq.(2) into the following right-hand sides,

$$\begin{aligned} n_{1y} &= \sum_{x' > 1, y' > y} \frac{4p_{x'y'}}{(x'-1)(y'-1)}, \\ n_{x1} &= \sum_{x' > x, y' > 1} \frac{4p_{x'y'}}{(x'-1)(y'-1)}, \\ n_{11} &= \sum_{x' > 1, y' > 1} \frac{4p_{x'y'}}{(x'-1)(y'-1)}, \end{aligned}$$

we obtain the distribution $P(A)$ of rectangles with the area A . The sum is taken over the positive integers $x', y' \leq L$, n_{11} , n_{x1} , and n_{1y} denote the numbers of the finally remaining rectangles with the areas 1×1 , $x \times 1$, and $1 \times y$, which are no more divisible. Note that the unknown factor p_{LL} disappears by the numerator and the denominator in all of $P(1) = n_{11}/\mathcal{N}$ and $P(x) = (n_{x1} + n_{1x})/\mathcal{N}$, where $\mathcal{N} = n_{11} + \sum_{x'' > 1} n_{x''1} + \sum_{y'' > 1} n_{1y''}$ denotes the total number of the divided rectangle faces.

Figure 2 shows the distribution of areas with width one. Our solution denoted by lines is almost completely fitting with the actual distribution denoted by marks. The part of linear tail becomes longer, as L is larger. Note that these distributions in Fig. 2 are estimated well by lognormal functions in this entire range of A .

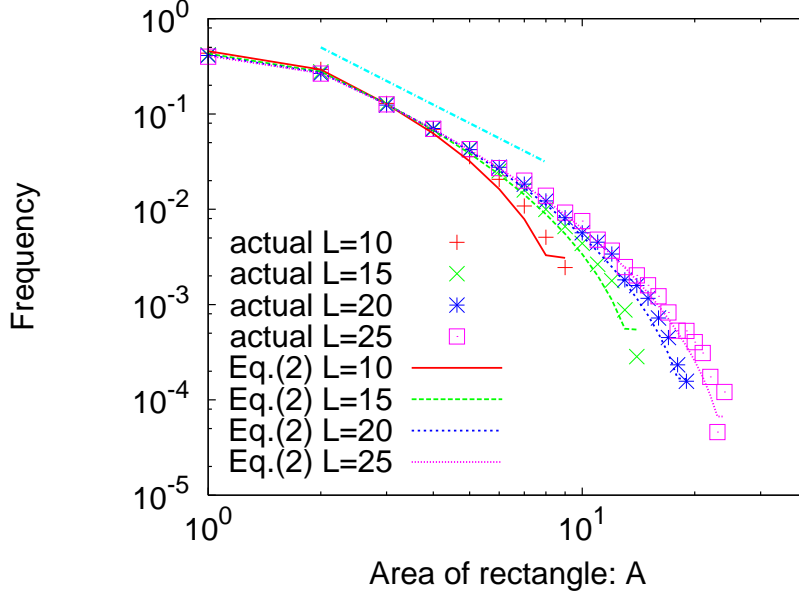


Fig. 2. (Color online) Distribution of areas: $1 \times 1, 1 \times 2, \dots, 1 \times L - 1$ in the extreme rectangles for $L = 10, 15, 20$, and 25 from left to right. The (plus, cross, star, and rectangle) marks show the averaged result in 100 samples of the actual divisions, and the corresponding (red, green, blue, and magenta) lines show the solution of Eq.(2) on the combinatorial analysis. The short (cyan) segment guides the slope of -2 corresponded to the exponent in broken fragments of glass [6] and city roads [8,9].

4 Continuous approximation

Since it possibly causes a combinatorial explosion to calculate the extreme distribution of areas: $1 \times 1, 1 \times 2, \dots, 1 \times L - 1$ at the absorbing states in the Markov chain, the application of Eq.(2) is restricted for a very small L . In order to analyze the distribution of areas for a large L , we approximate the process to be divisible at any positions on two edges of a rectangle, by ignoring the restriction to the segments on an $L \times L$ square lattice. For $l = 1, 2, \dots$ until the maximum layer at a given time, we consider the sum of the product of p_l and $g_{2l}(\log A)$, which denote the frequencies of layer l and of area A in the layer l . Here, the number l represents the depth of divisions. We derive these frequencies separately. Numerical simulations in subsection 4.3 show a good fitting of the mixture distribution $\sum_l p_l g_{2l}(\log A)$ with the actual distribution of areas in the average behavior of the divisions.

4.1 Distribution of layers

In this subsection, we ignore the difference of areas in each layer, and treat only the number of faces. Then, the stochastic division process characterized

as a Markov chain is simplified. Figure 3 shows the state transitions in the first few steps. We consider the number $n_l(t)$ of faces in the l -th layer at a time step t , and derive approximative solutions for the existing frequency of faces in the layer.

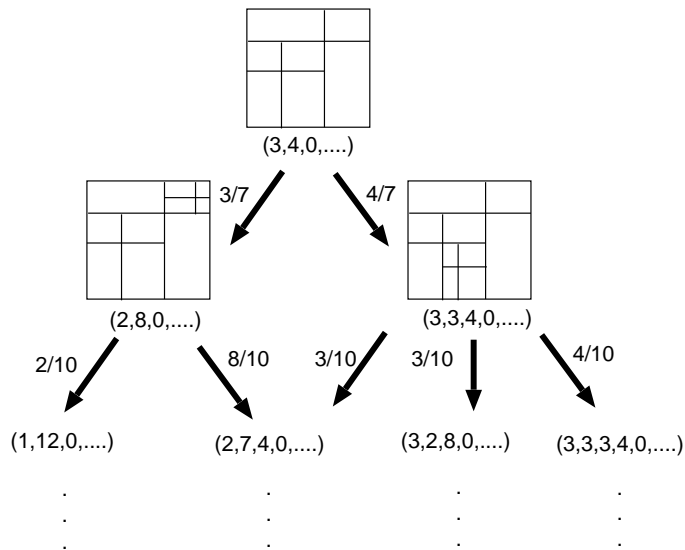


Fig. 3. Branching tree diagram of the state vector (n_1, n_2, \dots) for the division process at $t = 2, 3, 4$ steps (from top to bottom). Here, n_l is the number of faces on the l -th depth, when the difference of areas is ignored in the count. Each fraction denotes the transition probability. Note that two transitions until $t = 2$ through $(4, 0, \dots)$ at $t = 1$ and the initial square are trivial.

As shown in [23], the averaging behavior of difference

$$\Delta n_l \stackrel{\text{def}}{=} n_l(t+1) - n_l(t), \quad (3)$$

can be written to

$$\Delta n_l = m p_{l-1}(t) - p_l(t), \quad (4)$$

since a face in the layer l chosen with the probability $p_l(t)$ is divided into $m = 4$ smaller ones which belong to the layer $l + 1$, therefore a face in the layer $l - 1$ contributes to increase the number of faces in the layer l . Note that the simultaneously created m faces at a time belong to a same layer even with different areas. For a large t , by noticing $n_l(t) = \mathcal{N}(t)p_l(t)$ and substituting $\mathcal{N}(t) = \sum_l n_l = 1 + (m - 1)t \approx (m - 1)t$ into the right-hand side of Eq. (3), it is

$$\begin{aligned} \Delta n_l &= (m - 1)(t + 1)p_l(t + 1) - (m - 1)t p_l(t), \\ &= (m - 1)t[p_l(t + 1) - p_l(t)] + (m - 1)p_l(t + 1). \end{aligned}$$

Using $p_l(t+1) \approx p_l(t)$ because of $t+1 \approx t \gg 1$, Eq. (4) is rewritten to

$$p_l(t+1) - p_l(t) = -\frac{m}{(m-1)t} \{p_l(t) - p_{l-1}(t)\}, \quad (5)$$

where $p_0 \equiv 0$ is assumed for convenience. The solution of Eq.(5) is not easily derived even from a formal representation by using a generating function because of the combinatorial explosion involved with very complicated recursive operations [23]. For the $m = 2$ divisions, the difference equation (5) is equivalent to Eq.(14) in [19] for a crack model, however it differs to consider a rescaling length factor at each time step and to analyze a cumulative distribution in the one-dimensional case.

On the other hand, we also derive the following expression [23] by using a model in an interacting infinite particle system,

$$\frac{dn_l}{d\tau} = mn_{l-1} - n_l, \quad l \geq 2, \quad (6)$$

$$\frac{dn_1}{d\tau} = -n_1. \quad (7)$$

The solution is

$$n_l = \frac{(m\tau)^{l-1}}{(l-1)!} e^{-\tau}, \quad (8)$$

$$\mathcal{N}(\tau) = e^{(m-1)\tau}, \quad (9)$$

$$p_l = \frac{(m\tau)^{l-1}}{(l-1)!} e^{-m\tau}, \quad (10)$$

where $\tau \geq 0$ has a logarithmic timescale from the relation $1+(m-1)t = e^{(m-1)\tau}$ of the total number of faces. Note that the distribution of Eq.(10) coincides with the solution of Eq.(5) asymptotically after a huge time [23]. This form of n_l in Eq.(8) can be applied for calculating a fractal dimension at a proper time in the MSQ network based on a self-similar tiling, and extended to a preference model for selecting a face in Appendixes A and B.

4.2 Distribution of areas in the l -th layer

Remember that, at each time step, a rectangle is chosen uniformly at random (u.a.r). For the division, vertical and horizontal axes are also chosen u.a.r from the segments on an $L \times L$ square lattice. We focus on a set of rectangle faces

on only the l -th layer generated in some steps. After l -time selections, the subdivided face belongs to the layer l . The area S_l is given by the product of shrinking rates $0 < X_i, Y_i < 1$, $i = 1, 2, \dots, l$, for two edges of rectangle,

$$S_l = \prod_{i=1}^l X_i Y_i L^2,$$

where X_i and Y_i are rational numbers, and S_l is a positive integer in the division process, strictly speaking.

As an approximation for a large L , we assume that the random variables X_i and Y_i follow a $(0, 1)$ uniform distribution. Then we define a variable $x \stackrel{\text{def}}{=} -\log(S_l/L^2) = -\sum_i (\log X_i + \log Y_i)$ in the range of equivalent relation $x \geq 0 \Leftrightarrow L^2 \geq S_l$, the probability of x follows a gamma distribution

$$g_{2l}(x) = e^{-x} \frac{x^{2l-1}}{(2l-1)!}, \quad (11)$$

Here, the mean μ and the variance σ^2 of $\log X_i$ are

$$\mu = \int_0^1 \log X dX = -1,$$

$$\sigma^2 = \int_0^1 (\log X - \mu)^2 dX = 1.$$

Therefore, by a central-limit theorem, we have a normal distribution $N(0, 1)$ asymptotically

$$\frac{\log S_l - (\log L^2 - 2l)}{\sigma \sqrt{l}} \rightarrow N(0, 1),$$

$$\log L^2 - 2l = \log \left(\frac{L}{e^l} \right)^2.$$

In other words, the average shrinking rate of edge is $1/e$ for each division of a rectangle, and it is slightly smaller than $1/2$ for that of a square in the MSQ network. We can easily transform $g_{2l}(x)$ to the function of $\log S_l$ by using the shift of $x = 2 \log L - \log S_l$ for a constant L .

4.3 Simulation for the mixture distribution

In numerical simulation, we investigate the distributions decomposed into the approximative p_l and g_{2l} , and discuss the condition for a good fitting to each of components in the actual distributions for the divisions of rectangles. Indeed, our approximation of $\sum_l p_l g_{2l}$ shows reasonable agreement with the actual distribution of areas. Here, we consider the complementary cumulative distribution (CCD) of areas with a merit of smoothing effect, because the frequency distribution itself has a huge variety of areas especially for a large L , therefore it is practically impossible to gather these samples in a proper frequency.

Figure 4 shows the distribution p_l of layers at time steps $t = 50, 500$, and $5000 \ll T_{max} \sim L^{1.91}$ for $L = 10^8$ and 10^5 . By the effect of width one, there exists a gap between the solution of difference equation(5) denoted by open marks and the actual distribution denoted by lines, although these distributions almost coincide in the MSQ networks based on a self-similar tiling [23]. The gap becomes slightly larger as L is smaller in Fig. 4(b) and t is larger further to the right, because the effect tends to appear in more coarse-grained divisions and a deeper layer. With the growing of the total number $\mathcal{N} = 1 + (m - 1)t$ of faces, the depth of a face tends to be deeper as the time step t is larger. In addition, we note the expectation $\langle l \rangle \sim \log t$ [23]. The Poisson distribution of Eq.(10) denoted by closed marks has a slightly larger gap than the solution of difference equation(5) denoted by open marks. In the semi-log plots of Fig. 4(c)(d), we also investigate the fitting of the tails. In the tails, the discrepancy between the actual distribution and our approximation appears for the large $L = 10^8$ as t is larger further to the right, while these distributions do not fit any longer for the small $L = 10^5$. Here, in order to keep the accuracy of approximation, we set the initial condition of $\{p_l(0)\}$ by the existing frequency of faces in each layer at $t = 5$ that is explicitly determined from the branching tree diagram with the transition probability as shown in Fig. 3. Note that the calculation becomes more complex as t is larger, although a higher accuracy is expected.

Figure 5 shows the CCD of $x = \log(L^2/A)$ restricted in the l -th layer. We chose the most observable layer $l = 5, 8, 11$, and 14 , which correspond to the peaks of p_l at $t = 50, 500, 5000$, and $50000 \ll T_{max}$, respectively. Because the most observable layer is dominant in the mixture distribution $\sum_l p_l g_{2l}$. The effect of width one tends to appear as t is larger as shown in more right curves. The actual distributions denoted by lines and the gamma distribution $g_{2l}(x)$ of Eq.(11) denoted by marks almost coincide until increasing around $t = 5000$ step (the 3rd curve from left) in Fig. 5(a) for $L = 10^8$, however the distributions begin to differ from larger than $t = 500$ step (the 2nd curve from left) in Fig. 5(b) for $L = 10^5$. Thus, the effect of width one becomes stronger, as L is smaller, the discrepancy between the actual distribution and

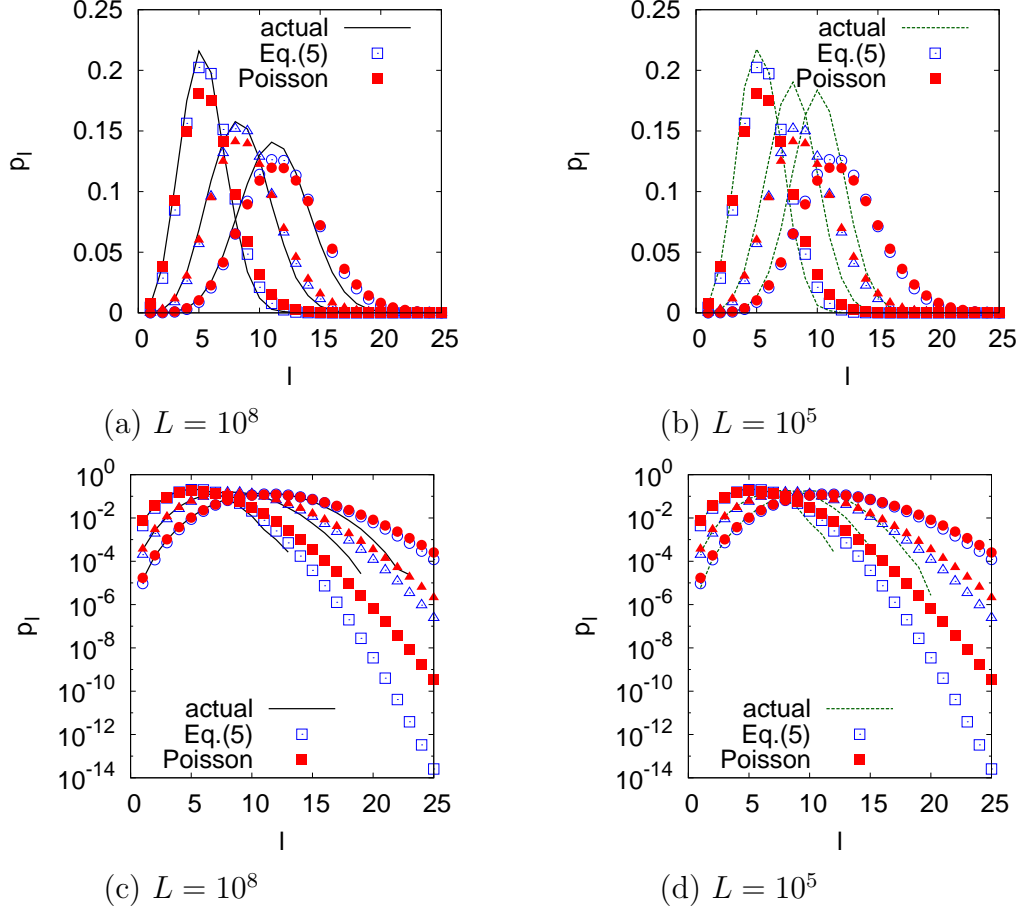


Fig. 4. (Color online) Distributions p_l of layers at time steps $t = 50, 500$, and 5000 from left to right. The (black) solid and (green) dashed lines correspond to the averaged result by 100 samples of the actual divisions of rectangles for (a) $L = 10^8$ and (b) $L = 10^5$, respectively. The semi-log plots are shown in (c) and (d) for investigating the fitting in the part of tail. The open and closed marks correspond to the solution of Eq.(5) and Poisson distribution in Eq.(10). The discrepancy of positions between the open marks and the lines is due to the effect of width one in the extreme rectangles. Note that these marks for $L = 10^8$ and 10^5 are the same at each time step, and that only the solid and dashed lines are different in (a) and (b) or (c) and (d).

the approximative g_{2l} is unignorable. We also confirm this phenomenon for the distributions in a commonly existing layer as shown in Fig. 6. The small cases of $L = 10^5$ and 10^6 give inaccurate approximations even at $t = 500$. In other word, for a smaller L at more coarse-grained divisions, the effect of width one already appears before $l = 12$ in a smaller (shallower) layer. This result is consistent with the average shrinking rate $1/e$ of edge per division, $e^{12} > 10^5$ and $e^{14} > 10^6$, as mentioned in subsection 4.2.

In order to grasp the tendency of approximation as a whole, we investigate the maximum difference in the CCD of areas between the actual data in the

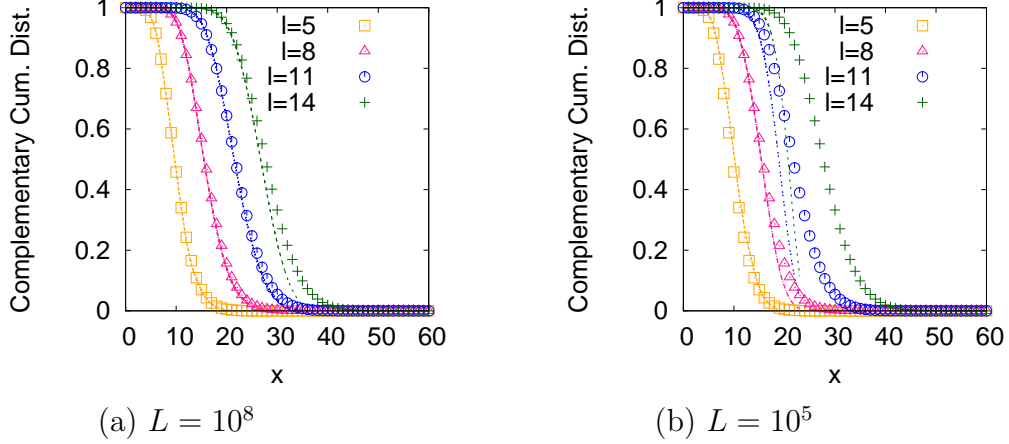


Fig. 5. (Color online) CCDs of $x \stackrel{\text{def}}{=} \log(L^2/A)$ restricted in the layer l at time steps $t = 50, 500, 5000$, and 50000 from left to right, which correspond to $l = 5, 8, 11$, and 14 at the peaks of p_l (see Fig. 4) in this order. The (orange, red, blue, and green) lines show the averaged results for 100 samples of the actual divisions of rectangles, while the corresponding marks (square, triangle, circle, and plus from left to right) show the results for the gamma distribution of Eq.(11).

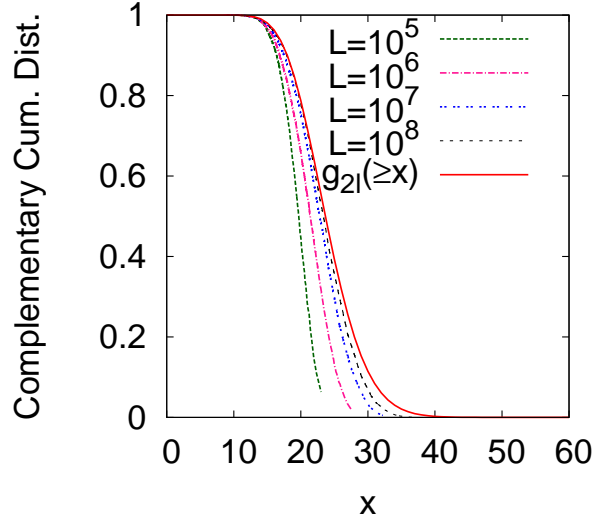


Fig. 6. (Color online) Comparison of the CCDs of x in a commonly existing layer $l = 12$ to the sizes of L at a time step $t = 500$. From left to right, the (green, magenta, blue, and black) dashed lines show the averaged results for 100 samples of the actual divisions of rectangles for $L = 10^5, 10^6, 10^7$, and 10^8 . The furthest right (red) solid line shows the result for the gamma distribution of Eq.(11).

divisions of rectangles and our approximation for several time steps. Figure 7 shows that the difference increases as t is larger and L is smaller. In each size of L , the lines with open marks give slightly better approximation than the lines with closed marks, where the former use the solution of difference equation (5) and the latter use the Poisson distribution (10) as p_l . The accuracy depends

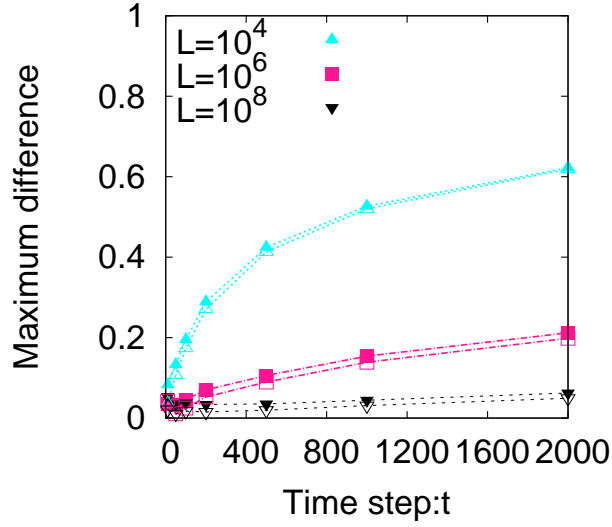
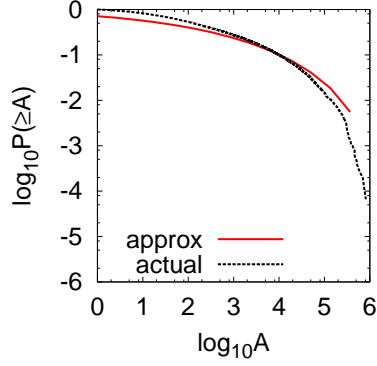


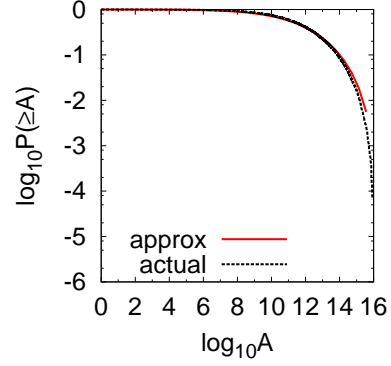
Fig. 7. (Color online) Maximum difference in the CCDs $P(\geq A)$ between the averaged result by 100 samples of the actual divisions of rectangles and our approximation by the mixture distribution $\sum_l p_l g_{2l}$ at several time steps t . From bottom to top, the (black, magenta, and cyan) lines with marks show the results for $L = 10^8, 10^6$, and 10^4 . The open and closed marks correspond to the mixtures with the solution of Eq.(5) and with the Poisson distribution of Eq.(10), respectively.

on the p_l as shown in Fig. 4 at a small t , while the effect of g_{2l} is added at a large t as shown in Fig. 5. The saturated behavior in Fig. 7 is not important, because the difference between any CCDs is bounded in $[0, 1]$ by nature.

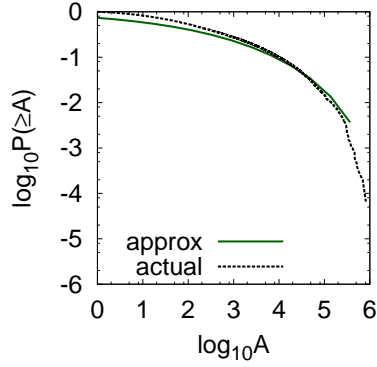
In more detail, we investigate the CCDs of areas at $t = 50$ step in Fig. 8. We obtain a good fitting for $L = 10^8$ in Fig. 8(b)(d), but remark on a small gap for $L = 10^3$ in Fig. 8(a)(c). Here, $L = 10^8$ and $L = 10^3$ are selected as examples to compare the accuracy of approximation affected by the appearance of width one in rectangles. The discrepancy between the actual distribution and our approximation at $t = 500$ step in Fig. 9(a)(c) becomes larger than the corresponding results at $t = 50$ in Fig. 8(a)(c) for $L = 10^3$, while the two distributions almost coincide in both Figs. 8 and 9 (b)(d) for $L = 10^8$. Figures 8(e)(f) and 9(e)(f) show reasonable fittings of the actual distribution with the estimated lognormal functions in the cumulation for the body over the whole range of $\log_{10} A$ and the tail restricted in the range of (e) $\log_{10} A \geq 5$ and (f) $\log_{10} A \geq 14$. For the body and the tail, Figure 10 shows the lognormal density functions. The short segment represents the slope of the exponent -2 for fragment-mass of glass [6] and areas enclosed by the city roads [8][9]. The linear part in Fig. 10 seems to be longer for larger L further to the right, as similar to the area distributions of extreme rectangles in Fig. 2. Remember that a large L gives fine-grained divisions. This phenomenon may be consistent with the enhancement of power-law property as high impact in the fragmentation of glass [6], however the behavior in our model is not exactly the same because of



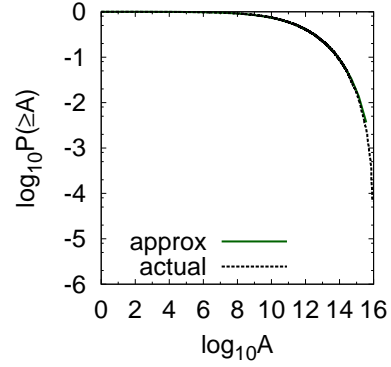
(a) Poisson, $L = 10^3$



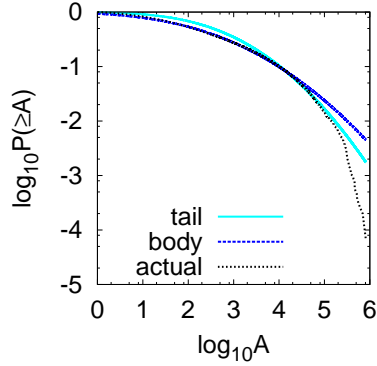
(b) Poisson, $L = 10^8$



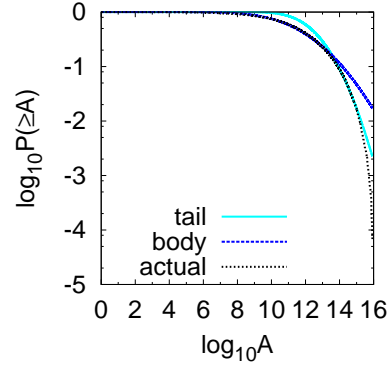
(c) Diff.Eq., $L = 10^3$



(d) Diff.Eq., $L = 10^8$



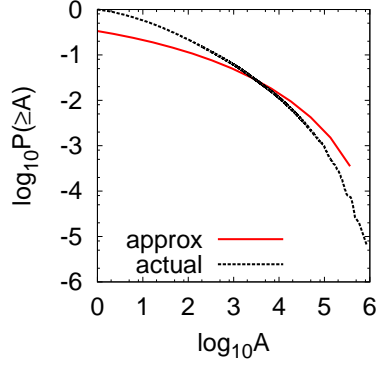
(e) lognormal, $L = 10^3$



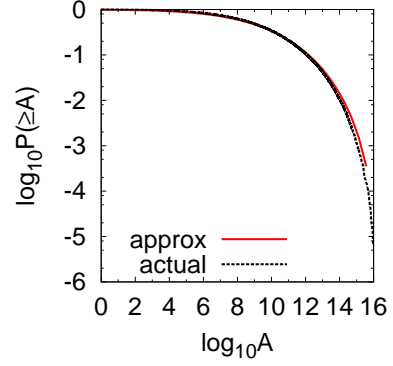
(f) lognormal, $L = 10^8$

Fig. 8. (Color online) CCDs $P(\geq A)$ at $t = 50$ step. The (red or green) solid and (black) dashed lines show our approximation by using (a)(b) the Poisson distribution of Eq.(10) or (c)(d) the solution of difference equation(5) and the averaged results by 100 samples of the actual divisions of rectangles, respectively. The (blue and cyan) lines show (e)(f) CCDs of the estimated lognormal functions for the body over the whole range of $\log_{10} A$ and the tail in (e) $\log_{10} A \geq 5$ and (f) $\log_{10} A \geq 14$.

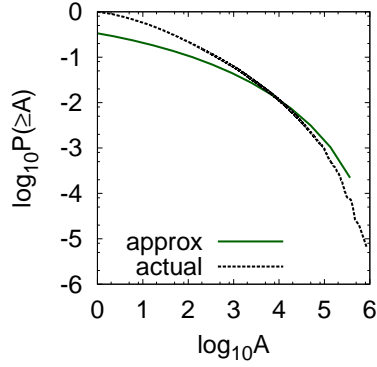
the remaining lognormal property whose distribution rather resembles to the ones for road networks in German cities [10]. We note that the approximation



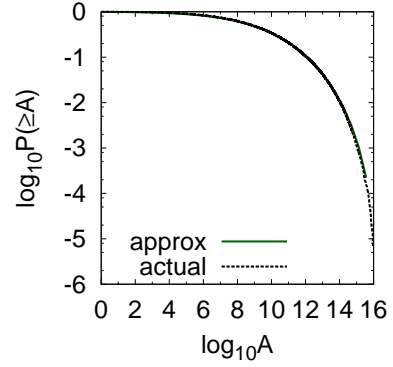
(a) Poisson, $L = 10^3$



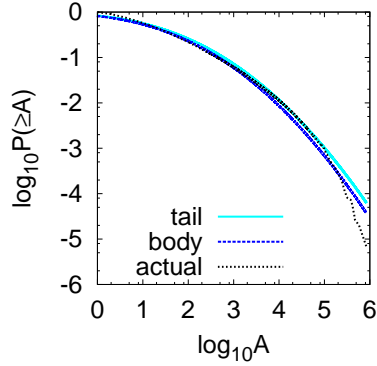
(b) Poisson, $L = 10^8$



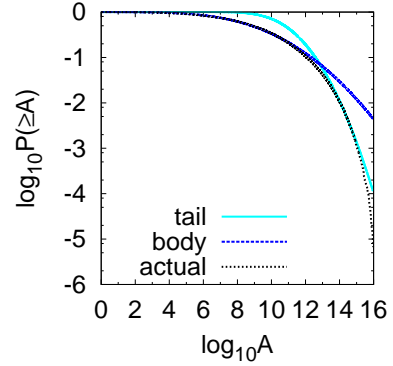
(c) Diff.Eq., $L = 10^3$



(d) Diff.Eq., $L = 10^8$



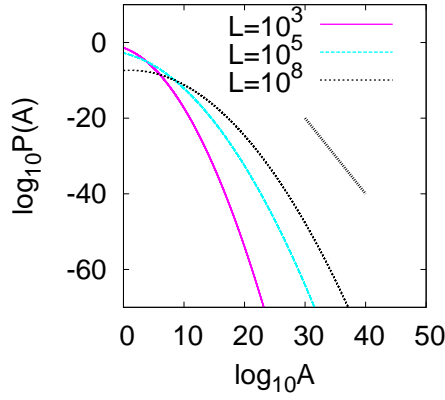
(e) lognormal, $L = 10^3$



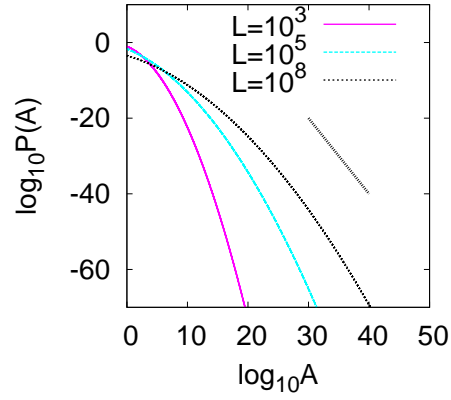
(f) lognormal, $L = 10^8$

Fig. 9. (Color online) CCDs $P(\geq A)$ at $t = 500$ step. Approximation by the mixture $\sum_l p_l g_{2l}$ using (a)(b) Poisson distribution of Eq.(10) or (c)(d) the solution of differential Eq.(5) as p_l , and (e)(f) the estimated lognormal functions for the body over the whole range of $\log_{10} A$ and the tail in (e) $\log_{10} A \geq 5$ and (f) $\log_{10} A \geq 14$. The linetypes are the same as in Fig. 8.

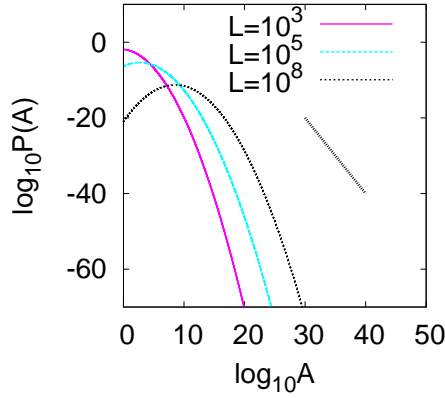
is no longer accurate for $L = 10^8$ in $t \gg 500$. The maximum difference < 0.1 in Fig. 7 may be a criterion for whether it gives a good fitting or not.



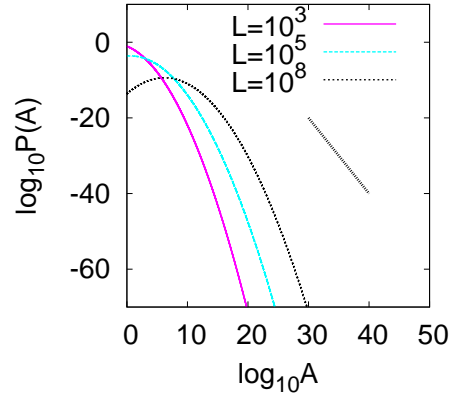
(a) Body, $t = 50$



(b) Body, $t = 500$



(c) Tail, $t = 50$



(d) Tail, $t = 500$

Fig. 10. (Color online) Approximation of area distributions estimated by lognormal functions in Fig. 8(e)(f) and Fig. 9(e)(f). The short segment guides the slope of -2.

5 Conclusion

Fat-tail distributions are pervasive in nature, and also appear in spatial networks. In particular, lognormal and power-law distributions are familiar in fragments of glass [6,7], crack patterns, and areas enclosed by city roads [8,9,10]. Beyond the details of physical phenomena, it is useful to consider a common generation mechanism on the division process. Thus, we have investigated a simple model for producing such distributions of areas enclosed by edges in a spatial network, which is an extension of MSQ networks [21,22,23] from the iterative divisions of squares to ones of rectangles.

The stochastic division process makes a Markov chain in the random selections of a rectangle face and of the division axes from the initial configuration of an $L \times L$ square. We have derived the exact solution of distribution at T_{max} for the extreme rectangles with no more divisible edge(s) of width one

on a combinatorial analysis for a small L . It is also pointed out that, in the absorbing Markov chain, the iterative divisions of rectangles are equivalent to directional random walks with splitting into four particles. For a large L and $t \ll T_{max}$, we have discussed the distribution of areas on a continuous approximation. As the distributions of layers and of areas restricted in a layer, we decompose the original distribution into two functions and consider the mixture of them, which corresponds to the Poisson and the gamma distributions in Eqs. (10) and (11). In addition, we obtain the distribution of layers by the difference equation (5) more accurately. Simulation results show a good agreement of our approximation with the actual distribution in the divisions, and give a condition for the fitting. We obtain more accurate results, as the size L is larger and the time step t is smaller, since the layer of a face tends to be shallower and the effect of width one in rectangles becomes weaker. We emphasize that the decomposition into two distributions of layers and of areas restricted in a layer will be useful for investigating other phenomena, such as in broken fragments of glass and areas enclosed by city roads, with additional information of layers defined by the time sequence of divisions.

We also confirm a slightly better agreement of our approximation for the $m = 2$ divisions which correspond to a crack model, though the meaning of network construction by bridgings should be discussed further from an application point of view. Unfortunately, relations to the mathematical properties of STIT tessellations [15][16][17] are unknown. In addition, there remain further challenges to the analysis for other properties, e.g., the lifetime of face or the distribution of edge lengths in our framework of stochastic processes, more rigorous investigation for the fitting functions (e.g., estimation by a double Pareto [3][5]), extension to a preference selection model, and considering the division by any direction not limited to vertical and horizontal.

Acknowledgment

The authors would like to thank Prof. Mitsugu Matsushita (Chuo University) for suggesting a similarity with the property for fragmentation of glass in Refs. [6] and [7]. This research is supported in part by a Grant-in-Aid for Scientific Research in Japan, No. 21500072.

Appendix A: Fractal dimension

We consider the fractal dimension d_f of a MSQ network [22] based on the self-similar tiling by squares at a finite time $\tau < \infty$. Note that the asymptotical value is $d_f \rightarrow 2$ trivially for the infinite time $\tau \rightarrow \infty$, since the division process

has no limitation and the whole two-dimensional space is embedded by the recursive subdivisions after a very long time.

In general, for the number $N_b[l]$ of covered boxes by a measure $\epsilon = 2^{-l}$, it is defined as

$$d_f = \lim_{l \rightarrow \infty} \frac{\log N_b[l]}{\log(1/\epsilon)} = \lim_{l \rightarrow \infty} \frac{\log N_b[l]}{l \cdot \log 2}.$$

For each visible level, the number $N_b[l]$ is counted as

$$N_b[2] = 2 \times N_b[1] + n_2(\tau),$$

$$N_b[3] = 2 \times N_b[2] + n_3(\tau),$$

:

where, in the above right-hand sides, the first term is due to the doubling of measure, and the second term comes from the one-to-one correspondence between four faces and cross edges generated by the quadratic division (e.g., in a clockwise mapping). The recursion is $N_b[l] = 2^{l-1}N_b[1] + \sum_{i=2}^l 2^{l-i}n_i(\tau)$ with $N_b[1] = 12$ in an initial configuration of four squares. By substituting Eq.(8) with $m = 4$ into $\sum_{i=2}^l 2^{l-i}n_i(\tau)$, we derive

$$\begin{aligned} \sum_{i=2}^l 2^{l-i} 4^{i-1} \frac{\tau^{i-1}}{(i-1)!} e^{-\tau} &= 2^l \sum_{i=2}^l \frac{(2\tau)^{i-1}}{2(i-1)!} e^{-\tau}, \\ &= 2^l (e^{2\tau} - 1 - Res) e^{-\tau} / 2, \end{aligned}$$

where Res denotes the residual for the higher-terms than l in the Taylor expansion of e^x .

For $l \gg 1$, we obtain

$$\log N_b[l] \sim \log(2^l(6 + e^\tau/2)) \sim l \cdot \log 2 + \tau. \quad (12)$$

On the other hand, by using a generating function $F_t(z) \stackrel{\text{def}}{=} \sum_{w \in Leaves_t} z^{|w|}$, it is also represented as

$$N_b[l] = 4 \times \sum_{w \in Leaves_t} 2^{l-|w|} = 4 \times 2^l F_t(1/2),$$

where $|w|$ denotes the depth of face w , and $Leaves_t$ is a set of faces at the division of t -step in the random quadtree, which corresponds to our hierarchical

network model. From the LEMMA 7.1 in [20], the expectation is

$$\mathcal{E}(N_b[l]) = 4 \times 2^l \mathcal{E}(F_t(1/2)),$$

$$\mathcal{E}(F_t(z)) \leq \exp \left((4z - 1) \sum_{k=0}^{t-1} \frac{1}{3k+1} \right).$$

When we set $3k+1 = x$,

$$\mathcal{E}(F_t(1/2)) \leq \exp \left(\int_1^{3t-2} \frac{1}{3x} dx \right) \leq (3t+1)^{1/3}.$$

From the relation $1+3t = e^{3\tau}$ of the total number of faces, we obtain

$$\log \mathcal{E}(N_b[l]) \leq \log 4 + l \cdot \log 2 + \frac{1}{3} \log(1+3t) \sim l \cdot \log 2 + \tau.$$

This is equivalent to Eq.(12).

We consider a typical time $\langle \tau \rangle = \sum_{\tau} \tau p_l(\tau) \times m$ for the layer l ,

$$\langle \tau \rangle \approx m \int_0^{\infty} \frac{m^{l-1}}{(l-1)!} \tau^l e^{-m\tau} d\tau = \frac{l}{m},$$

where the factor m is due to the normalization,

$$\sum_{\tau} p_l(\tau) \approx \frac{1}{(l-1)!} \int_0^{\infty} (m\tau)^{l-1} e^{-m\tau} d\tau = \frac{1}{m}.$$

Thus, as an upper-bound for a large l , we obtain

$$d_f \sim \frac{l \cdot \log 2 + l/m}{l \cdot \log 2} = 1 + \frac{1}{m \cdot \log 2} = 1.36067.$$

Note that this value is between Koch curve: $\log 4 / \log 3 = 1.26186$ and Sierpinski carpet: $\log 8 / \log 3 = 1.89278$ or Sierpinski gasket: $\log 3 / \log 2 = 1.58496$.

Appendix B: Extended preference model

When a rectangle face is chosen (not uniformly at random but) proportionally to the power γ^l of its depth l with a real parameter $\gamma > 0$, we can approxi-

mately derive the distribution $p_l = n_l/\mathcal{N}$. Note that, in the selected division, a larger face (at a shallower layer) is preferred for $\gamma < 1$, while a smaller face (at a deeper layer) is preferred for $\gamma > 1$ [20].

The number n_l of l -th faces at time τ follows

$$\frac{dn_l}{d\tau} = m \frac{\gamma^{l-1}}{C} n_{l-1} - \frac{\gamma^l}{C} n_l, \quad l \geq 2, \quad (13)$$

$$\frac{dn_1}{d\tau} = -\frac{\gamma^1}{C} n_1. \quad (14)$$

As $\gamma \approx 1$, we assume that $1/C$ is a constant. The reason is discussed later. By a variable transformation $\tau = C\gamma^{-l}T$, we rewrite Eq.(13) as

$$\frac{\gamma^l}{C} \frac{dn_l}{dT} = \frac{dn_l}{d\tau} = \frac{\gamma^l}{C} \left(\frac{m}{\gamma} n_{l-1} - n_l \right).$$

Thus, we obtain the same form of Eq.(6) at $\gamma = 1$ as follows

$$\frac{dn_l}{dT} = \frac{m}{\gamma} n_{l-1} - n_l.$$

The solution is

$$n_l = \left(\frac{m}{\gamma} \right)^{l-1} \frac{T^{l-1}}{(l-1)!} e^{-T} = \left(\frac{m}{\gamma} \right)^{l-1} \frac{(\gamma^l \tau / C)^{l-1}}{(l-1)!} e^{-\gamma^l \tau / C},$$

$$\mathcal{N} = \sum_{l=1} n_l = \sum_{l=1} \frac{(m\gamma^{l-1}\tau/C)^{l-1}}{(l-1)!} e^{-\gamma^l \tau / C}.$$

Here, we evaluate the assumption of a constant $1/C$. Since the number of faces increases by $m-1 = 3$ (add four divisions, and delete one chosen face) at each time step t , the total number of faces is $\sum n_l = 1 + (m-1)t$. It must be equal to Eq(9), so that $\sum n_l = e^{(m-1)\tau}$. By applying $d\sum n_l/d\tau = (m-1)\sum n_l$, the following relation must hold from Eqs.(13) and (14),

$$\sum_{l=1} \frac{dn_l}{d\tau} = \frac{m-1}{C} \sum_{l=1} \gamma^l n_l = (m-1)e^{(m-1)\tau},$$

therefore $C = (\sum_l \gamma^l n_l) e^{-(m-1)\tau}$. Indeed, around $\gamma \approx 1$, $1/C$ is almost constant as approximated in Table 1.

τ	1	2	3	4
$\gamma = 0.95$	1.21	1.42	1.66	1.93
$\gamma = 0.97$	1.12	1.25	1.38	1.52
$\gamma = 0.99$	1.04	1.08	1.12	1.16
$\gamma = 1.01$	0.96	0.92	0.88	0.84
$\gamma = 1.03$	0.88	0.75	0.61	0.52
$\gamma = 1.05$	0.77	0.46	0.34	0.29

Table 1

The estimated values of $1/C$ by the Newton-Raphson method for varying a parameter γ . Note that τ takes a small value even for a huge network because τ is a logarithmic timescale for a linear timescale $t = 1, 2, \dots$ of discrete steps in the relation $1 + (m - 1)t = e^{(m-1)\tau}$.

References

- [1] H.A. Simon, *Biometrika*, **42**, 425–440, (1955).
- [2] M.E.J. Newman, A.-L. Barabási, and D.J. Watts, Chapter Three: Empirical Studies, pp.167–228, *The Structure and Dynamics of NETWORKS*, Princeton University Press, 2006.
- [3] M. Mitzenmacher, *Internet Mathematics*, **1(2)**, 226–251, (2004).
- [4] M. Mitzenmacher, *Internet Mathematics*, **1(3)**, 305–333, (2004).
- [5] W.J. Reed, and M. Jorgensen, *Communications in Statistics: Theory and Methods*, **33(8)**, 1733–1753, (2004).
- [6] H. Katsuragi, D. Sugino, and H. Honjo, *Phys. Rev. E* **70**, 065130, (2004).
- [7] T. Ishii, and M. Matsushita, *J. of The Physical Society of Japan*, **61(10)**, 3474–3477, (1992).
- [8] S. Lämmer, B. Gehlsen, D. Helbing, *Physica A* **363**, 89, (2006).
- [9] A.P. Masucci, D. Smith, and C.M. Batty, *Eur. Phys. J. B* **71(2)**, 259, (2009).
- [10] S.H.Y. Chan, R.V. Donner, and S. Lämmer, *Eur. Phys. J. B* **84(4)**, 563–577, (2011).
- [11] G.W. Delaney, S. Hutzler, and T. Aste, *Phys. Rev. Lett.* **101**, 120602, (2008).
- [12] P.S. Dodds, and J.S. Weitz, *Phys. Rev. E* **67**, 016117, (2003).
- [13] M. Barthelemy, and A. Flammini, *Phys. Rev. Lett.* **100**, 138702, (2008).
- [14] S.-H. Lee, and P. Holme, arXiv:1205.0537, (2012).
- [15] W. Nagel, and V. Weiss, *Adv. Appl. Prob.(SGSA)*, **37**, 859–883, (2005).

- [16] W. Nagel, J. Mecke, J. Ohser, and V. Weiss, *Image Anal. Stereol.*, **27**, 73–78, (2008).
- [17] C. Thäle, *Image Anal. Stereol.*, **28**, 69–76, (2009).
- [18] V. Kalapala, V. Sanwalani, A. Clauset, and C. Moore, *Phys. Rev. E* **73**, 026130, (2006).
- [19] R. Cowan, *Advances in Applied Probability* **42(1)**, 26–47, (2010).
- [20] D. Eisenstat, *Proceeding of SIAM the 8th Workshop on Analytic Algorithms and Combinatorics* (ANALCO11), Jan. 22, 2011.
<http://arxiv.org/abs/1008.4916>
http://www.siam.org/proceedings/analco/2011/anl11_09_eisenstatd.pdf
- [21] Y. Hayashi, *Physica A* **388**, 991, (2009).
- [22] Y. Hayashi, and Y. Ono, *Phys. Rev. E* **82**, 016108, (2010).
- [23] Y. Hayashi, *IEICE Trans. Fundamentals*, **E94-A(2)**, 846-849, (2011).
- [24] M.I. Karavelas, and L.J. Guibas, *Proc. of the 12th ACM-SIAM Symposium on Discrete Algorithms*, (2001).
- [25] T.M. Liggett, *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*, Springer, 1999.
- [26] B. Derrida, M.R. Evans, V. Hakim, and V. Pasquier, *J. Phys. A* **26**, 1493, (1993).